

Article

Disenchanted Trust: Instrumental Reason, Algorithmic Governance, and China's Emerging Social Credit System

Sheng Zou

International Institute, University of Michigan, Ann Arbor, MI 48109, USA; E-Mail: szou@umich.edu

Submitted: 30 October 2020 | Accepted: 25 January 2021 | Published: 6 April 2021

Abstract

Digital technologies have provided governments across the world with new tools of political and social control. The development of algorithmic governance in China is particularly alarming, where plans have been released to develop a digital Social Credit System (SCS). Still in an exploratory stage, the SCS, as a collection of national and local pilots, is framed officially as an all-encompassing project aimed at building trust in society through the regulation of both economic and social behaviors. Grounded in the case of China's SCS, this article interrogates the application of algorithmic rating to expanding areas of everyday life through the lens of the Frankfurt School's critique of instrumental reason. It explores how the SCS reduces the moral and relational dimension of trust in social interactions, and how algorithmic technologies, thriving on a moral economy characterized by impersonality, impede the formation of trust and trustworthiness as moral virtues. The algorithmic rationality underlying the SCS undermines the ontology of relational trust, forecloses its transformative power, and disrupts social and civic interactions that are non-instrumental in nature. Re-reading and extending the Frankfurt School's theorization on reason and the technological society, especially the works of Horkheimer, Marcuse, and Habermas, this article reflects on the limitations of algorithmic technologies in social governance. A Critical Theory perspective awakens us to the importance of human reflexivity on the use and circumscription of algorithmic rating systems.

Keywords

algorithmic rationality; Frankfurt School; instrumental reason; Social Credit System; social governance; trust

Issue

This article is part of the issue "Critical Theory in a Digital Media Age: Ways Forward" edited by Robert E. Gutsche, Jr. (Lancaster University, UK).

© 2021 by the author; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

The development of big data and algorithmic technologies has enabled governments across the world to fashion new modes of political and social control. An epitome of this emerging trend of algorithmic governance is China's plan to build a Social Credit System (SCS), which has evoked fear internationally of an Orwellian technodystopia. The system is intended to aggregate data on both natural and legal persons in order to monitor, evaluate, and modify their actions through a joint mechanism of reward and punishment. Instead of seeing the SCS as an exclusive symbol of Chinese authoritarianism, we should situate it in a global context of algorithmic governmentality while recognizing its embeddedness in local

political and cultural traditions. The use of algorithmic analysis in governmental practice is not unique to China, but is in place in Western countries as well, especially with respect to policing and criminal justice (e.g., Angwin, Larson, Mattu, & Kirchner, 2016; Dencik, Hintz, Redden, & Warne, 2018; Richardson, Schultz, & Crawford, 2019). A close look at the China case will inform larger discussion of algorithmic governance across the world. Framed by the Chinese government as an endeavor to build trust in society, the SCS represents the colonization of the everyday life by ascendant logics of quantification, measurability, and efficiency—or in short, the quantification of the social (Mau, 2019). But can trust be built through algorithmic quantification and top-down schemes of governance aimed to nudge, constrain, and manipulate

human behavior into compliance? Although a growing literature has empirically investigated the mechanics of the SCS in China, this fundamental theoretical question remains unanswered.

This article addresses this void and interrogates, more broadly, the increasing embrace of algorithmic rationality in social governance. I first sketch out the current shape of the SCS in China in relation to evolving discourses around it, and propose a conception of the SCS as a project of moral engineering with a focus on trust-building. Then, I centralize and delve into the moral and relational dimension of trust in social interactions. In the rest of the article, I draw on the Frankfurt theorists' critique of formalized or instrumental reason, particularly the works of Horkheimer, Marcuse, and Habermas, to elaborate on the ways in which the SCS—as an epitome of the quantification of the social—disenchants and flattens moral values such as trust and trustworthiness.

2. Beyond Surveillance: Social Credit System as Moral Engineering

China's SCS has attracted global attention since 2014, after the Party-state released the Planning Outline for the Establishment of a Social Credit System ("Establishment of a social credit system," 2015), which laid out goals to put the system in place by 2020. A simple Google search of China's SCS returns links to Western media coverage that compares it to the dystopian world depicted in the *Black Mirror* episode, "Nosedive" (Schur, Jones, & Wright, 2016), where people rate each other for every interaction, which impacts their socioeconomic statuses. A growing amount of research in this area has debunked these reductionist caricatures (e.g., Ahmed, 2019; Creemers, 2018; Ohlberg, Ahmed, & Lang, 2017), and has shown that, far from an established all-encompassing system that assigns everyone a single score, the current state of the SCS consists of dozens of government-led pilot projects at local or national levels and various commercial ones run by tech giants such as Alibaba and Tencent. At the national/central level, the National Development and Reform Commission, the Supreme People's Court, and the People's Bank of China, among others, have been taking the lead in initiating nationwide SCS pilots, including the formation of an interministerial joint conference, a national financial credit information database, and a portal called Credit China that publicizes SCS policies as well as blacklists and red lists. At the municipal and county levels, local governments have been experimenting with their own SCS systems, many of which are based on quantified scoring. Although these programs are led by the government, the collaboration between governmental and commercial actors have also been noteworthy, with the latter providing strong technological and infrastructural support. While commercial SCS experiments are more akin to loyalty programs (Creemers, 2018), those run by the government are more invasive and wider in scope.

One of the most notorious examples of government-led pilots is the case of Rongcheng, a county-level city in Shandong Province, where residents are assigned scores on a scale of 1,000 and classified into descending levels from A to D. The evaluation system covers a range of behaviors and activities: economic, social, civic, and moral. Misdemeanors, such as jaywalking, littering, and getting traffic tickets, lead to score deduction and punishment, while exemplary behaviors, such as caring for elderly parents, helping others, donating to charity, and volunteering for public programs, translate into score bumps and benefits (Mistreanu, 2018; Ohlberg et al., 2017). Benefits may come in the form of deposit waivers for bike rental, discounts on heating bills, or advantageous terms on bank loans (Mistreanu, 2018), while penalties include limited access to government benefits and restrictions on market entry (Creemers, 2018).

Admittedly, not all SCS pilots utilize quantified schemes; some renowned ones simply take the forms of blacklists and red lists, such as the judgement debtors list administered by the Chinese Supreme People's Court. However, there has been growing reliance on quantification across local trials. Liu (2019) documents that by mid-2019, 21 Chinese cities had enacted their own quantified SCS pilots, and 27 more cities were in the preparatory stage. Notable cases include Fuzhou's Jasmine (Moli) score that rates citizens on a 0–1,000 scale, and Suzhou's Osmanthus (Guihua) score that assigns citizens up to 200 points.

The SCS represents the Party-state's latest endeavor to modernize and automate its social governance. The root of this cybernetic mode of control can be traced all the way back to the 1970s and 1980s, when the Party leadership and intelligentsia began discussing the automation of social governance. The cybernetic imagination advanced by Qian Xuesen, China's Father of Rocketry, and Song Jian, a cybernetics expert, shaped former president Hu Jintao's concept of scientific development, where scientific and engineering approaches were imagined as solutions to problems in the social domain (Hoffman, 2017). In the early 1990s, the idea of building a credit system was broached in response to Chinese firms' debt default problem; in 2002, then-President Jiang Zemin voiced the demand for a SCS to regulate market behavior (Liang, Das, & Kostyuk, 2018). Early efforts in building the credit system focused primarily on the financial sector; the Planning Outline released in 2014, however, significantly extended the scope of the project to the social domain.

Western media coverage and some scholarship on the Chinese SCS heavily focus on the issue of surveillance, privacy, and political control; Liang et al. (2018), for instance, frame this project as a 'state surveillance infrastructure.' Yet surveillance is not the end, but a means through which the authorities bring the citizens' behaviors—potentially political but mostly civic—in line with the official ideologies and values. The 2014 Planning Outline clearly indicates the moral education component

of the SCS, which is aimed to cultivate a culture of trust and integrity. The SCS, in this sense, fits into the Party's long tradition of 'spiritual civilization' campaigns. Hence, I propose an alternative framework of the SCS as a project of moral engineering with emphasis on trust-building.

Some scholars, such as Lee (2014), characterize China as a low-trust society, where distrust of strangers is prevalent largely because the society is traditionally structured around familial or kinship ties. Moreover, the informal networks of trust were undermined through incidents of political turmoil, such as the Cultural Revolution. Post-reform China witnessed rapid economic growth along with unethical get-rich-quick schemes and the corruption of moral integrity. A slew of social issues has become disheartening over the past decades, such as food safety, corruption, fraud, tax evasion, incivility etc. Officially, the SCS has been framed as a mechanism for building social trust and a culture of integrity; it covers a range of relationships among the state, the corporate sector, and individual citizens. The 2014 Planning Outline lays out the goal to build a joint mechanism of reward and punishment, which "mak[es] it so that the trustworthy benefit at every turn and the untrustworthy can't move an inch" ("Establishment of a social credit system," 2015). Local pilots have also significantly focused on the cultivation of trustworthiness and civility, such as the Rongcheng case and, more recently, the controversial and short-lived Suzhou Civility Code pilot that encouraged citizens to sort their trash, follow traffic rules, and engage in volunteer services (Chiu, 2020). In this sense, the SCS is an extension of the Party-state's lasting scheme to cultivate its citizens' *suzhi* (human quality).

Public opinion in China has appeared to embrace the role of the SCS to enhance trust and civility in society. A nationwide survey on Chinese citizens' perception of the SCS revealed high levels of approval: 80% of the 2,209 respondents either somewhat or strongly approved the system (Kostka, 2019). Privacy infringement is not the dominant frame people use to interpret the SCS; rather, the initiative is often considered conducive to promoting honest dealings in society and economy (Kostka, 2019).

In this sense, the SCS can be construed as a project of moral engineering. Pre-digital or non-digital strategies in this regard include promoting model citizens, enacting the *dang'an* system with dossiers kept on individuals' trajectories (Jiang, 2020), and promulgating spiritual civilization campaigns such as 'Eight Virtues and Eight Shames.' Of course, the SCS, with the use of digital technologies, differs greatly from its antecedents in terms of its workings and underlying logic. Much of the extant literature on the SCS either tiptoes around or only scratches the surface of the issue of trust. This article seeks to fill this void by returning to basic questions on the ontology and ethics of trust in relation to the technologization of social governance.

3. Trust as a Moral Concept: Grounding Social Trust in Moral Philosophy

3.1. Trust as a Floating Signifier in Official Planning of the Social Credit System

The term 'trust' is recurring in the 2014 Planning Outline on the construction of the SCS. In some sense, it can be read as the authorities' *modus operandi* to utilize auratic terms such as 'trust' as a justificatory cover for political and ideological control. The notion of 'trust' is thus reduced to what Marcuse (1964/2002) calls a self-validating 'magic-ritual' formula hammered and re-hammered into people's minds—a mere governing device that precludes the development of meaning. There may be some level of sincerity in the authorities' intention to reshape the moral landscape of the society, but whenever trust is invoked, its content is taken for granted as a given that entails no explication.

In fact, trust is an exceedingly nebulous concept that can be defined from different perspectives. What complicates things even further is that the all-encompassing and slippery word for 'trust' in Chinese (*xin*) can, in other contexts, mean 'credit,' 'integrity,' 'confidence,' 'belief,' or 'faith.' There has been no clear analytic distinction made between these synonyms in the official discourse around the SCS. In fact, it is exactly by invoking the ambiguous notion of *xin* that the authorities manage to conflate the notion of trust with its numerous synonyms, thus quietly extending the scope of its control from market behavior to social and civic conduct. The 2014 Planning Outline envisions the SCS to be a panacea for problems in four different domains: administrative and government affairs, business and commercial activities, judicial affairs, and social interactions. In the document, the Chinese words *zhengxin* (financial credit) and *chengxin* (trustworthiness or integrity) are both used under the overarching rubric of 'social credit.' Moreover, input data in the design of local pilots often include variables from different sectors—administrative, judicial, economic, and social. While overdue loans or legal violations may lead to low scores, such deeds as donating blood and volunteering would be rewarded in local pilots conducted in cities such as Xiamen and Fuzhou (Lewis, 2019). The rewards and punishment also apply to a wide range of scenarios that impact people's life, such as housing, employment, medical care, and public services, among others. In one extreme case, a Chinese dating website boosted visibility of users with higher Sesame credit score—a commercial scoring system created by Alibaba—placing their profiles in prominent spots (Hatton, 2015).

What sets the SCS apart from Western credit scoring systems is the state's attempt to regulate not only economic activities but also social behaviors, and the overt top-down scheme of reward and punishment stretching into private domains of everyday life. In this process, trust becomes a floating signifier across different

sectors and contexts, whose meanings seem to be self-evident yet never clearly delineated. 'Trustworthiness' (as in social interactions) is conflated with 'creditworthiness' (as in market transactions) or even compliance with regulations (as in civil and judicial domains). Indifferent to the specific workings and meanings of trust in different contexts, the SCS is largely built on the rationale of pre-existing quantified ratings of creditworthiness. Yet is it sensible to quantify and rate trustworthiness, a moral virtue, based on a FICO-score-like model? This is a basic question too readily brushed aside.

Admittedly, even the financial credit rating *per se* is not free from associations with moral concepts such as honesty and integrity (Lauer, 2017); credit scores, when taken far beyond their original turf and used as proxies of other virtues in hiring and promotion processes, can yield damaging effects on one's life (O'Neil, 2016). Nonetheless, credit scoring is widely accepted in many circumstances due to their efficiency in regulating economic and transactional activities within certain rule-bound contexts, typically marketplaces. Compared to the moralization of creditworthiness, the SCS's logic of quantifying morality is equally, if not more, pernicious as it, by design, applies a marketplace-based governmentality to the non-economic realm—particularly the social, the civil, and the interpersonal. Trust in marketplace and business settings is oftentimes calculative—and perhaps rightfully so—because transactions within this context are heavily instrumental, whereas trust in wider social (and non-economic) interactions is relational, as it is anchored in "social relationships when there are strong beliefs about the goodwill, honesty, and good faith efforts of others" (Poppo, Zhou, & Li, 2016, p. 724). The muddling of calculative and relational dimensions of trust amounts to the confusion between what Habermas (1984) would call instrumental/strategic action and communicative action. While strategic action strives for influence and instrumental ends, communicative action seeks to reach genuine understanding (Habermas, 1984). To Habermas, the issue with technological modernization lies in the encroachment of technical efficiency in the realm of communicative action (Pippin, 1995). The urge to quantify trust in social interactions is reflective of this trend. Yet to see trust as a calculable matter fails to do justice to the rich moral connotations of relational trust, that is, a general belief in the integrity of others during social interactions.

3.2. Trust and Moral Autonomy

To ground relational trust in moral philosophy, I follow Seligman's (1997) comparison of trust and its synonym 'confidence.' In part informed by Luhmann (1988), Seligman (1997) posits that confidence hinges on one's perception of the reliability of a social system and its enforcement of role expectations; in other words, confidence is placed upon institutional authority, such as religion and kinship in traditional societies, and the market

or the state in modern societies (Lee, 2014). In contrast, trust emerges in response to "the breakup of primordial forms of social organization and the greater differentiation of social roles," which results in more intersubjective negotiation and higher contingency (Lee, 2014, p. 5). To put it in a nutshell, when confidence in a system of social role expectations can no longer be taken for granted—and when the possibility of dissonance in role fulfilment arises—there emerges the need for trust as a form of social relations (Seligman, 1997). Seligman, following Luhmann, thus locates trust in the realm of interpersonal and social connections, or in the encounter between self and alter, recognizing the moral autonomy of both.

Moreover, Seligman stresses that trust, different from confidence, implies risk and uncertainty. Trust is incurred often when "the acts, character, or intentions of the other cannot be confirmed" (Seligman, 1997, p. 21). While trust involves uncertainty and vulnerability, confidence often pertains to the reliability of the other's words, commitments, or acts based on past knowledge and future possibilities of sanctions (Seligman, 1997). Trust, in this sense, also differs from a contract. A contract explicitly lays out the agreed-upon terms and rules binding the involved parties, with the threat of sanctions, whereas trust operates upon the freedom from such formal agreement and deterrence. That is not to say that there is no obligation to be trustworthy, but this obligation "arises from moral agency and autonomy, from the freedom and responsibility, of the participants to the interaction" (Seligman, 1997, p. 6).

Trust in social interactions should then be construed not as an inert substance, but as a relational practice that involves moral agency and autonomy. To take Seligman's framework further, I echo Flores and Solomon (1998) in locating trust within a "dynamic emotional relationship which entails responsibility...[and] a set of social practices, defined by our choices, to trust or not to trust" (p. 205). This kind of trust carries a moral valence.

Both trusting others and being trustworthy are moral virtues, because trust-giving requires the trustor's benevolence to prevail over the unpredictability of others' intentions and over potential risks, and being trustworthy entails the trustee's responsibility to fulfill promises and expectations as well as to reciprocate the trustor's kindness. Trust thus subsumes risks, enabling interactions that would otherwise be impossible. Of course, as a choice, to trust or not to trust is never guaranteed, and may involve some level of calculation. But trust in social interactions is never a mere function of such calculation. A person may choose to trust someone despite the low confidence in the latter's ability to reciprocate. In the worst-case scenario, the trust relationship breaks down eventually; and the best outcome is that the trustee, moved by the trustor's sincerity, manages to honor the trust relationship against all odds. Trust as such never preempts any possible trajectory and gives space to the moral agency of both parties involved.

As moral virtues, trust and trustworthiness in the realm of communicative action naturally resist quantification and datafication. The impulse to quantify trustworthiness through algorithms—and to use such quantification as the basis of trust-giving—is symptomatic of the deepening technologization of society and instrumentalization of reason. To critically examine the ascendant trend to quantify moral worth, we may revisit and extend the Frankfurt School scholars' trenchant critique of the technological society and formalistic rationality in mid-20th century. Informed by their theories, I discuss in the rest of the article how the algorithmic rationality (Lowrie, 2017) or algorithmic governmentality (Rouvroy, 2013) animating the SCS contradicts the ontology of relational trust and undermines its formation in social/civic interactions.

4. A Critical Theory Perspective on the Social Credit System and Trust-Building

4.1. From Instrumental Reason to Algorithmic Rationality

One of the main concerns of the first-generation Frankfurt theorists, especially Horkheimer and Marcuse, is the interplay between technology, rationality, and domination. They take issue with the ascendancy of late capitalism and techno-science as the twin forces of domination, which undergird a type of instrumental reason. Their discussion of instrumental reason and unreflective technical process is largely informed by Weber's (1921/1978) distinction between formal rationality and substantive rationality (Gunderson, 2015). Weber (1919/1946) traces in modern bureaucracies the displacement of traditional and value-laden understanding of the world with a technical and impersonal mode of thinking and behaving, oriented towards efficiency and instrumental goals; this type of formal-rational understanding disenchant the lifeworld and erodes traditional values. Weber's theory influenced Horkheimer's (1947/2004) notion of instrumental reason and Marcuse's (1964/2002) discussion of technological rationality.

Horkheimer (1947/2004) refers to the formal rationality as subjective reason/rationality. In *Eclipse of Reason*, he deals at length with the process whereby objective reason, characteristic of the pre-modern era, has been replaced by subjective reason in modern societies. With objective reason, one judges actions and objects as good or bad, according to their harmony (or the lack thereof) with the objective world as a reasonable system. The subjective reason, however, signifies the formalization and instrumentalization of reason and manifests itself in one's "ability to calculate probabilities and thereby to coordinate the right means with a given end" (Horkheimer, 1947/2004, p. 4). Subjective reason is formalized in that it is purged of moral and aesthetic reflections and concerned merely with "the adequacy of procedures for purposes more or less taken for granted" (Horkheimer,

1947/2004, p. 3), thus precluding deliberation on the meaning and merit of substantive goals; it is instrumentalized in that it is aimed only to attain subjective interests or self-preservation—the utmost 'reasonable' goal in modern societies. Subjective reason is incapable of evaluating the quality of an object or action in terms of good or bad, and is only interested in their utility.

Frankfurt theorists associate this type of instrumental rationality to modern science and technology, which are firmly grounded in positivism and naïve empiricism. As Horkheimer (1947/2004, p. 53) contends, the positivists reduce science to its empirical procedures, make it "absolute as truth," and rely on the scientific successes to justify their methods. Positivism separates fact from value and, when extended to the social domain, leads to the reification of life and perception and the erosion of human agency.

In the same vein, Marcuse (1964/2002) discusses how the operational thinking of positivist science treats any concept as no more than a set of operations; and how, on a societal level, the metaphysical conception of subjectivity is replaced by what he calls 'a one-dimensional society,' where people's thoughts and behaviors are molded in a way that conforms to the dominant power in society with the aid of unreflective technological progress. To Marcuse (1964/2002), science and technology in late capitalist societies are not neutral, because, as a socio-historical project, they operate in a given universe of discourse and action, and are already structured in a particular way. Domination in modern society, he argues, not only perpetuates itself *through* technology but *as* technology, in the sense that technology at once legitimates and glosses over the dominant power with its aura of objectivity. Both Horkheimer and Marcuse point to the paradox where rationality progresses into irrationality.

Central to the workings of modern science and technology is the logic of quantification and calculability. Not only is there a latent link among quantification/calculation, knowability, and domination in theory, but in praxis, there has been a long history in which bureaucrats and technocrats have deployed statistical and mathematical tools for social governance in early modern societies (e.g., Desrosières, 1998; Foucault, 2007; Porter, 1996). With the increasing mechanization of governance embraced by bureaucratic institutions, numbers have been relied on as a vital means of classification and control.

Frankfurt theorists trenchantly critique the expansion of this logic of quantification and calculability. Obsession with calculability, according to Horkheimer (1947/2004), supplants meanings with function or effect. The only 'reasonable' words are those that can be technically calculated for possibilities. Moral judgment about good or bad, in their unverifiable forms, are deemed useless. To make these evaluative criteria 'reasonable,' scientific operationalization is entailed; that is how the incalculable enters into the scientific horizon—"through a series of reductions" (Marcuse, 1964/2002, p. 142).

The quantification of qualities, Marcuse (1964/2002) argues, constitutes a particular way of seeing, anticipating, and projecting, one that separates the reality from all inherent ends and paves the way for domination. Under the logic of quantification, values and qualities “lose their mysterious and uncontrollable character” and “appear as calculable manifestations of (scientific) rationality” (Marcuse, 1964/2002, p. 172): They are mutilated and reduced to their behavioral translation.

The algorithmic rationality that animates the SCS is an extension of the instrumental or technological rationality in Frankfurt theorists’ critiques. Nonetheless, algorithmic rationality also differs from traditional scientific rationality and signifies a novel epistemic paradigm. While the sine qua non of traditional scientific rationality has been the work of proof (Lakatos, 1976), the goal of data-driven algorithms is not to prove anything, but to achieve feasibility, practicality, and efficiency (Lowrie, 2017). Algorithmic processes bypass the systems of verification associated with traditional science that “usually appear essential to attest to the robustness, truth, validity or legitimacy of claims and hypotheses formulated about reality” (Rouvroy, 2013, p. 151). Algorithmic analysis is not about causes, but rather about correlations and probabilities based on patterns of data.

The workings of algorithms are undergirded by a type of anticipatory rationality (Amoore, 2013; Gillespie, 2014; Hong & Szpunar, 2019), aimed at predicting and preempting possible futures. It “separates the individuals from their possibility of not actualizing what they are capable [of], or their possibility of not being subjected to what they are likely to be subjected to” (Rouvroy & Stiegler, 2016, pp. 10–11). In this way, the preemptive logic acts on the present while also shaping the future, actualizing some potentialities while usurping others. Yet this process is smoothed over by the cold charisma of digital technologies that conjures up an aura of objectivity and impartiality (boyd & Crawford, 2012; Gillespie, 2014). Although algorithms are often interconnected with human experts, they are also expected to act on their own and thus possess their own agency and authority. This is reminiscent of Horkheimer’s (1947/2004, p. 87) diagnosis that “the machine has dropped the driver.”

The Frankfurt School offers a broader historical framework for the reflection of algorithmic rationality that progresses from older forms of scientific and technological rationality. It shows that the progression of technological rationality is not inevitable but historically contingent, shaped by shifting configurations of political and economic interests. What is hinted at but underexplored in their theorization is that such a shift to instrumental reason is as philosophical and epistemological as it is moral. The rejection of an objective moral order itself presupposes a certain view on morality or a certain moral economy—to borrow Daston’s (1995) term, namely a system of values carrying affective and moral valences. Objectivity and impartiality are part and parcel of the moral economy on which instrumental rea-

son thrives. In the following, I shall extend the Frankfurt School’s theory on instrumental reason through the lens of moral economy, and apply it to the case of quantified SCS pilots. I argue that algorithmic scoring of trustworthiness defeats the purpose of moral engineering as it undermines the ontology of relational trust by authorizing a moral economy antithetical to the workings of trust in social and civic relationships.

4.2. Disenchanting Trust: The Paradox of Moral Engineering

As an embodiment of the formalization of reason, the quantified SCS pilots seeks to transmute trustworthiness into a calculative and knowable matter represented by standardized numerical values. Admittedly, in certain rule-bound contexts such as a marketplace, technologies of quantification have some merits in mediating exchanges and transactions. Porter (1996), for instance, argues that quantification as “a technology of distance” is suited for communication extending beyond the boundaries of locality and community (p. ix); it facilitates intellectual exchanges and economic transactions since numbers render “dissimilar desires, needs, and expectations” commensurable (p. 86). It is worth noting that Porter’s focus rests on the tension between human expertise and mechanical objectivity in professional and scientific fields (e.g., actuarial science, accounting, engineering, medicine) instead of intersubjective encounters in the private domain. In fact, technologies of quantification were developed to replace personal trust in situations where institutional authority was weak (Daston, 1995; Porter, 1996). In this sense, the rise of technologies of quantification is symptomatic of the declining personal trust rather than a solution to restoring it. Likewise, the quantified SCS systems in China could be understood as a symptom of the declining moral authority of the communist ideology.

In the present day, while quantification of trustworthiness in transactional relationships (e.g., rating of vendors on Amazon and ratings of drivers on Uber) minimizes risks and transaction costs, its application to trust formation in non-instrumental interactions is problematic. The all-encompassing design of the SCS ignores such contextual differences. Notwithstanding its moralizing framing, the SCS as a project of moral engineering is self-defeating for various reasons. Since oft-mentioned issues of privacy infringement, opacity, inaccuracy, and bias have been thoroughly explored in the works on politics of algorithms, they will not be belabored here. Instead, my focus is placed upon how quantified SCS systems impede the formation of relational trust.

First, as discussed above, relational trust is best construed as a moral virtue at the discretion of individuals, but the quantified SCS pilots contradict its ontology. The technological quantification of trust relies on a moral economy of impersonality and impartiality, particularly “the reining in of judgment, the submission to rules, the reduction of meanings” (Daston, 1995, p. 19). It is

precisely a moral economy inhospitable to the formation of relational trust, which entails the exercise, rather than the withholding, of judgment. In the moral economy characterized by impersonality and impartiality, individuals are treated as mere technical objects rather than “ethical or aesthetic or political agent[s]” (Marcuse, 1964/2002, p. 150). They become ‘one-dimensional’ not only in Marcuse’s sense of the word—as their critical consciousness is whittled away—but also because they are collapsed into one-dimensional digital profiles, into digits and numbers. Accordingly, such systems shift attention from human relationships to technical efficacy, and thereby relegate trust to mere confidence in technologies. I have discussed above the distinction between trust and confidence, and how the official framing of the SCS strategically muddles this distinction. The confusion of these two concepts is, in some sense, a result of the increasing technologization of society at large. O’Reilly (2013), for instance, touts the benefits of algorithmic regulation, including the affordance of data-driven reputation systems, which he believes could “improve the outcomes for citizens with less effort by overworked regulators” (p. 294). Technologies are imagined as solutions for social problems. This type of ‘solutionism’ (Morozow, 2014) resonates with the elitist cybernetic thinking in China and with the project of the SCS. It appears sensible that issues of trust in a technological society would be readily resolved through technological means. However, as Nissenbaum (1999) cautions, the attempt to attain trust through technical means (e.g., cybersecurity measures) is misguided, since when we seek to attain security, we are usurping the role of trust. She cites evidence showing that the extrinsic constraints, such as surveillance, could ultimately diminish trust because intrinsic motivation is stifled.

The reduction of trust to confidence in technologies prompts tactical maneuvers such as gaming. Instead of empowering individuals to build and exercise trust, the scoring systems nudge them to gain more points. Individuals only need to come up with whatever methods to increase the score—or to coordinate input variables with expected output. Whether these methods are morally good or not becomes irrelevant. Individuals no longer need to interact with others to determine trustworthiness and the quality of trust relationships; rather, they are prompted to interact more frequently with algorithms to monitor their own standings. This transference of evaluation from human reason to algorithmic rationality is a contemporary manifestation of the formalized/instrumental reason, a type of rationality only concerned with the coordination of effective means with taken-for-granted ends.

Second, the technological quantification of trust, by actualizing certain possibilities while preempting others based on cost-benefit analysis, forecloses the transformative power of trust. Trust is precious in social relationships because it subsumes risks and enables human interactions that would be otherwise impossi-

ble. Although it does not preclude the possible breakdown of a given relationship, it opens up possibilities for the enhancement of human bonds and the nourishment of moral characters. With the development of cold and authoritative algorithms, such an unrealized space of possibilities is converted into a verifiable temporal sequence (Reigeluth, 2014). Through the reduction of risk and uncertainty, algorithmic analysis of trustworthiness inhibits, rather than enables, risk-taking behavior and preempts its potentialities to generate favorable outcomes. Moreover, the unreflective use of data about individuals’ past behaviors to extrapolate their future trajectories unfairly assumes the permanence of their propensities while ignoring their potential to transform themselves for the better, hence the perpetuation of their undesirability. And it has been noted that the SCS pilots lack uniform mechanisms for credit repair for low scorers (Ohlberg et al., 2017). Some who commit misdemeanors may have undergone extenuating situations to which the one-dimensional scoring system is indifferent. They may be helped through others’ willingness to engage them instead of further punishment. The joint reward and punishment mechanism envisioned in the planning of the SCS risks duplicating punishment across a range of unrelated domains and perpetuating the vicious cycle that entraps low scorers. The preemptive logic of algorithmic analysis annuls the transformative agency of individuals in changing both others and themselves; it contradicts the very end of moral engineering, which is to cultivate subjects into better citizens.

Third, further disenchanting the moral concept of trust is the SCS’s conflation of economic and social behaviors and its confusion of financial creditworthiness with general trustworthiness. This tendency reflects a general process of the economization of society (Mau, 2019), where economic logics carry over into non-economic spheres. Intentionally extending the model of financial credit rating to non-economic spheres makes it so that ratings are no longer bounded within a particular marketplace, platform, or context; nor are they restricted to rational-economic interactions, but spill over indiscriminately to any spheres of life, giving rise to the reification of values. Trustworthiness is not only rendered calculable under the economic rationale, but is incentivized with material rewards as well. In Shanghai, for instance, a government-run social credit initiative offers trustworthy youths one-year access to rent-free apartments to reward their commitment to volunteer work (Yu, Liu, & He, 2018). Yet the true aura of trustworthiness and integrity as moral characters lies precisely in their independence from the economic and instrumental logics and from external incentives. The attempts to tether non-instrumental acts with material ends erode the intrinsic authority of moral norms.

Last but not least, in the case of SCS, algorithms that represent the will of the state loom above individuals in social interactions as a mighty third party, disrupting the trust relationship as a ‘moral party of two,’

to borrow Bauman's (1993) term (see also Lee, 2014). Admittedly, since the socialist era, the Party-state has sought to regulate the private life of citizens through various means and organizations (e.g., neighborhood committees). The use of quantified SCS systems to intervene in communal or even familial relationships as well as stranger sociality is an extension of this tradition. In the Rongcheng SCS pilot, for instance, deeds worthy of reward even include caring for elderly parents (Ohlberg et al., 2017). Mediated through algorithmic technologies, regulation of such private behavior operates quietly in the background as if bypassing the political authority. However, as the Frankfurt theorists caution, technologies are deeply connected with political power. The SCS seeks to bring individuals into compliance with the hegemonic values defined and promoted by the Party-state. Yet this technologically enhanced attempt to engineer morality and civility is self-defeating because it erodes the very foundation on which moral and civic norms rest.

Looking at the state-led 'spiritual civilization' campaigns in the 20th century China, Lee (2014) argues that the state management of social interactions, especially stranger sociality, intervenes in and disrupts the moral encounter where the self is supposed to assume moral responsibility in their direct dealings with others. Invoking Bauman's view (1993) on 'the moral party of two,' Lee (2014, p. 20) argues that a moral relation is "an affair of two, which cannot open up to a third party or an authority figure." Therefore, Lee (2014) argues that morality cannot be entrusted to the state or the market but should be cultivated via a robust civil society. Extending Lee's conception of moral relations as uncodifiable, I contend that trustworthiness and civility will not thrive under the weight of state-imposed and technologically mediated monitoring and nudging. Although trust, trustworthiness, and civility may be induced through top-down interventions such as the promulgation of norms and moral education (Luhmann, 1979; Nissenbaum, 1999), technologies alone will not solve the core issue of cultivating a social climate amenable to trust formation. To build a culture of integrity, space should be carved out for individuals to practice trust and build trust relations or civic bonds with others in everyday settings through trial and error, instead of falling back on a mighty third party for a confidence rating. Assuming the moral engineering project is sincerely meant to cultivate a more civilized citizenry, it ought to be, at the very least, decoupled from practices of law enforcement and market regulation, which are currently lumped together under the overarching rubric of the SCS.

5. Conclusions

This article examines China's emerging SCS as a project of moral engineering through the lens of Critical Theory, addressing the fundamental theoretical question as to whether the SCS could, as the official discourse proposes, provide a solution to the moral crisis in the rapidly

developing society. The quantified SCS pilots conflate economic and social behaviors and rely on algorithmic rationality for the evaluation of one's trustworthiness, disregarding the meanings and workings of trust and trustworthiness in different contexts (e.g., marketplaces vs. communities). The community-based relational trust is crudely reduced to calculative trust; efficiency and impersonality are privileged over subjective discretion and autonomy. Such tendencies are symptomatic of the increasing technologization of governance in modern societies. However, trust in social and civic interactions needs to be cultivated and practiced through one's encounters with others; and trustworthiness is not to be standardized and flattened into algorithmically generated numerical values.

Trust entails risk, uncertainty, responsibility, moral autonomy, and possibilities without guarantee, whereas these are precisely the elements that the algorithmic rationality is aimed to foreclose. To build trust, we need a different logic than instrumentality—a mode of thought that militates against naive operationalism and ubiquitous quantification (Marcuse, 1964/2002). We need to put trust back into moral relations between individuals, with their moral autonomy restored so that they can make evaluations with their individual reason, emotion, and judgment. It is misleading to believe that, in the case of trust-building, technological rationality is superior to traditional moral education, even though it may appear more efficient. In fact, it is precisely such efficiency that undermines the very cause of trust building.

Of course, it would be unhelpful and naïve to advocate a romantic return to the pre-modern era; nor is it the intention of the Frankfurt theorists. Instead, they are envisioning possibilities to reconstruct technologies in order to contribute to human freedom and possibilities. This requires a shift away from the unreflective positive thinking to a critical consciousness. Hence, instead of doing without technologies, we ought to reflect on when (and how) to incorporate technologies into human decision-making, and when (and how) to resist technological colonization of everyday life. We ought to contemplate the meanings of moral values as well as the quality of our goals rather than merely the efficacy of technological means. A critical rethinking of ethical issues around algorithms and governance through the case of China's emerging SCS also provides insights into existing practices of credit/reputation rating, user profiling, and algorithmic governance in Western societies. Re-reading Critical Theory in the digital age heightens our vigilance in the application of rating systems to expanding areas of everyday life, and our reflexivity on the use and circumscription of these systems.

Acknowledgments

The author would like to thank the anonymous reviewers for their elaborate, insightful, and constructive feedback on the previous version of this article.

Conflict of Interests

The author declares no conflict of interests.

References

- Ahmed, S. (2019, May 1). The messy truth about social credit. *Logic Magazine*. Retrieved from <https://logicmag.io/china/the-messy-truth-about-social-credit>
- Amoore, L. (2013). *The politics of possibility: Risk and security beyond probability*. Durham, NC: Duke University Press.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. Retrieved from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Bauman, Z. (1993). *Postmodern ethics*. Oxford: Blackwell.
- boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society*, 15(5), 662–679.
- Chiu, K. (2020, September 8). Suzhou city takes a page from China's social credit system with Civility Code that rates citizens' behaviour through a smart-phone app. *South China Morning Post*. Retrieved from <https://www.scmp.com/abacus/tech/article/3100516/suzhou-city-takes-page-chinas-social-credit-system-civility-code-rates>
- Creemers, R. (2018, May 22). *China's social credit system: An evolving practice of control*. SSRN. <http://dx.doi.org/10.2139/ssrn.3175792>
- Daston, L. (1995). The moral economy of science. *Osiris*, 10, 2–24.
- Dencik, L., Hintz, A., Redden, J., & Warne, H. (2018). *Data scores as governance: Investigating uses of citizen scoring in public services*. Cardiff: Data Justice Lab. Retrieved from <https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf>
- Desrosières, A. (1998). *The politics of large numbers: A history of statistical reasoning*. Cambridge, MA: Harvard University Press.
- Establishment of a social credit system. (2015, April 27). *China Law Translate*. Retrieved from <https://www.chinalawtranslate.com/en/socialcreditsystem>
- Flores, F., & Solomon, R. C. (1998). Creating trust. *Business Ethics Quarterly*, 8(2), 205–232.
- Foucault, M. (2007). *Security, territory, population: Lectures at the Collège de France*. New York, NY: Picador.
- Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality and society* (pp. 167–193). Cambridge, MA: MIT Press.
- Gunderson, R. (2015). Environmental sociology and the Frankfurt School 1: Reason and capital. *Environmental Sociology*, 1(3), 224–235.
- Habermas, J. (1984). *The theory of communicative action* (T. McCarthy, Trans.). Boston, MA: Beacon Press.
- Hatton, C. (2015, October 26). China's "social credit": Beijing sets up huge system. *BBC*. Retrieved from <https://www.bbc.com/news/world-asia-china-34592186>
- Hoffman, S. (2017). Programming China: The Communist Party's autonomic approach to managing state security. *MERICs China Monitor*, 44, 1–12.
- Hong, S., & Szpunar, P. M. (2019). The futures of anticipatory reason: Contingency and speculation in the sting operation. *Security Dialogue*, 50(4), 314–330.
- Horkheimer, M. (2004). *Eclipse of reason*. New York, NY: Continuum. (Original work published 1947)
- Jiang, M. (2020). A brief prehistory of China's social credit system. *Communication and the Public*, 5(3/4), 93–98.
- Kostka, G. (2019). China's social credit systems and public opinion: Explaining high levels of approval. *New Media & Society*, 21(7), 1565–1593.
- Lakatos, I. (1976). *Proofs and refutations*. Cambridge: Cambridge University Press.
- Lauer, J. (2017). *Creditworthy: A history of consumer surveillance and financial identity in America*. New York, NY: Columbia University Press.
- Lee, H. (2014). *The stranger and the Chinese moral imagination*. Stanford, CA: Stanford University Press.
- Lewis, D. (2019, October 11). All carrots and no sticks: A case study on social credit scores in Xiamen and Fuzhou. *Digital Asia Hub*. Retrieved from <https://www.digitaliasiahub.org/2019/10/11/all-carrots-and-no-sticks-a-case-study-on-social-credit-scores-in-xiamen-and-fuzhou>
- Liang, F., Das, V., & Kostyuk, N. (2018). Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy & Internet*, 10(4), 415–453.
- Liu, C. (2019). Multiple social credit systems in China. *Economic Sociology: The European Electronic Newsletter*, 21(1), 22–32.
- Lowrie, I. (2017). Algorithmic rationality: Epistemology and efficiency in the data sciences. *Big Data & Society*, 4(1), 1–13.
- Luhmann, N. (1979). *Trust and power*. New York, NY: John Wiley & Sons.
- Luhmann, N. (1988). Familiarity, confidence, trust: Problems and perspectives. In D. Gambetta (Ed.), *Trust: Making and breaking of cooperative relations* (pp. 94–107). Oxford: Basil Blackwell.
- Marcuse, H. (2002). *One-dimensional man*. London: Routledge. (Original work published 1964)
- Mau, S. (2019). *The metric society: On the quantification of the social*. Medford, MA: Polity Press.
- Mistreanu, S. (2018, April 3). Life inside China's social credit laboratory: The party's massive experiment in ranking and monitoring Chinese citizens has already started. *Foreign Policy*. Retrieved from <https://>

foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory

- Morozow, E. (2014, July 19). The rise of data and the death of politics. *The Guardian*. Retrieved from <http://www.theguardian.com/technology/2014/jul/20/rise-of-data-death-of-politics-evgeny-morozov-algorithmic-regulation>
- Nissenbaum, H. (1999). Can trust be secured online? A theoretical perspective. *Etica e Politica*, 1(2), 1–15.
- Ohlberg, M., Ahmed, S., & Lang, B. (2017). Central planning, local experiments: The complex implementation of China's social credit system. *MERICs China Monitor*, 43, 1–15.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, NY: Crown.
- O'Reilly, T. (2013). Open data and algorithmic regulation. In B. Goldstein & L. Dyson (Eds.), *Beyond transparency: Open data and the future of civic innovation* (pp. 289–300). San Francisco, CA: Code for America Press.
- Pippin, R. B. (1995). On the notion of technology as ideology. In A. Feenberg & A. Hannay (Eds.), *Technology and the politics of knowledge* (pp. 43–61). Bloomington, IN: Indiana University Press.
- Poppo, L., Zhou, K. Z., & Li, J. J. (2016). When can you trust "trust"? Calculative trust, relational trust, and supplier performance. *Strategic Management Journal*, 37(4), 724–741.
- Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- Reigeluth, T. (2014). Why data is not enough: Digital traces as control of self and self-control. *Surveillance & Society*, 12(2), 243–354.
- Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review*, 94, 192–233.
- Rouvroy, A. (2013). The end(s) of critique: Data-behaviourism vs. due-process. In M. Hildebrandt & K. De Vries (Eds.), *Privacy, due process and the computational turn: The philosophy of law meets the philosophy of technology* (pp. 143–168). Milton Park and New York, NY: Routledge.
- Rouvroy, A., & Stiegler, B. (2016). The digital regime of truth: From the algorithmic governmentality to a new rule of law. *La Deleuziana: Online Journal of Philosophy*, 3, 6–29.
- Schur, M., Jones, R. (Writers), & Wright, J. (Director). (2016). Nosedive [Television series episode]. In S. Brooker (Executive producer), *Black mirror*. Los Gatos, CA: Netflix.
- Seligman, A. B. (1997). *The problem of trust*. Princeton, NJ: Princeton University Press.
- Weber, M. (1946). Science as vocation. In H. H. Gerth & C. W. Mills (Eds.), *From Max Weber: Essays in sociology* (pp. 129–156). New York, NY: Oxford University Press. (Original work published 1919)
- Weber, M. (1978). *Economy and society: An outline of interpretive sociology*. Berkeley, CA: University of California Press. (Original work published 1921)
- Yu, R., Liu, J., & He, F. (2018, June 11). Chengxin you "jia" shouxin youyong [Integrity is "valuable," trustworthiness is useful]. *Xinhua Net*. Retrieved from http://www.xinhuanet.com/local/2018-06/11/c_1122967204.htm

About the Author



Sheng Zou is a Postdoctoral Research Fellow at the University of Michigan's International Institute. He received his PhD in Communication from Stanford University. His research interests include global media industries, digital media and society, platform economy and labor, and emerging technologies. His current work explores the intersection of governance, politics, and popular culture in digital China, particularly state-society interactions through new media platforms.